Summer Program FGV/EMAp 2019

Introduction to Machine Learning with Python

PRINCIPAL COMPONENT ANALYSIS

Prof. Luis Gustavo Nonato

University of São Paulo - São Carlos - SP

Principal Component Analysis

Principal Component Analysis

PCA is directly related to the eigenvectors and eigenvalues of covatiance matrices.

Principal Component Analysis

PCA is directly related to the eigenvectors and eigenvalues of covatiance matrices.

Lets so make a quick review of eigenvectors, eigenvalues, and covatiance matrices.

Eigenvectors and Eigenvalues

Eigenvectors and Eigenvalues

Given a $d \times d$ matrix $\mathbf{A}$, a pair $(\lambda, \mathbf{u})$ that satisfies

$$\mathbf{A}\mathbf{u} = \lambda\mathbf{u}$$

is called eigenvalue (λ) and corresponding eigenvector $(\mathbf{u})$ of $\mathbf{A}$.

Eigenvectors and Eigenvalues

Given a $d \times d$ matrix $\mathbf{A}$, a pair $(\lambda, \mathbf{u})$ that satisfies

$$\mathbf{A}\mathbf{u} = \lambda \mathbf{u}$$

is called eigenvalue (λ) and corresponding eigenvector ($\mathbf{u}$) of $\mathbf{A}$.

Symmetric Matrices

Eigenvectors and Eigenvalues

Given a $d \times d$ matrix $\mathbf{A}$, a pair $(\lambda, \mathbf{u})$ that satisfies

$$\mathbf{A}\mathbf{u} = \lambda\mathbf{u}$$

is called eigenvalue (λ) and corresponding eigenvector ($\mathbf{u}$) of $\mathbf{A}$.

Symmetric Matrices

- $\lambda \in \mathbb{R}$ and $\mathbf{u} \in \mathbb{R}^d$ (no complex numbers involved).

Eigenvectors and Eigenvalues

Given a $d \times d$ matrix $\mathbf{A}$, a pair $(\lambda, \mathbf{u})$ that satisfies

$$\mathbf{A}\mathbf{u} = \lambda \mathbf{u}$$

is called eigenvalue (λ) and corresponding eigenvector ($\mathbf{u}$) of $\mathbf{A}$.

Symmetric Matrices

- $\lambda \in \mathbb{R}$ and $\mathbf{u} \in \mathbb{R}^d$ (no complex numbers involved).
- The eigenvectors are orthogonal

$$\mathbf{u}_i^\top \mathbf{u}_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{otherwise} \end{cases}$$

(assuming $\|\mathbf{u}_i\| = 1$)

Symmetric Matrices

$\mathbf{A}$ symmetric with distinct eigenvalues λ_i.

The equations $\mathbf{A}\mathbf{u}_i = \lambda_i \mathbf{u}_i$ can be written in matrix form as:

Symmetric Matrices

A symmetric with distinct eigenvalues λ_i.
The equations $\mathbf{A}\mathbf{u}_i = \lambda_i \mathbf{u}_i$ can be written in matrix form as:

$$\underbrace{\begin{bmatrix} a_{11} & & a_{1d} \\ \vdots & \dots & \vdots \\ a_{d1} & & a_{dd} \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_d \\ | & & | \end{bmatrix}}_{\mathbf{U}} = \underbrace{\begin{bmatrix} | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_d \\ | & & | \end{bmatrix}}_{\mathbf{U}} \underbrace{\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_d \end{bmatrix}}_{\mathbf{D}}$$

Symmetric Matrices

$\mathbf{A}$ symmetric with distinct eigenvalues λ_i.
The equations $\mathbf{A}\mathbf{u}_i = \lambda_i \mathbf{u}_i$ can be written in matrix form as:

$$\underbrace{\begin{bmatrix} a_{11} & & a_{1d} \\ \vdots & \dots & \vdots \\ a_{d1} & & a_{dd} \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_d \\ | & & | \end{bmatrix}}_{\mathbf{U}} = \underbrace{\begin{bmatrix} | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_d \\ | & & | \end{bmatrix}}_{\mathbf{U}} \underbrace{\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_d \end{bmatrix}}_{\mathbf{D}}$$

In matrix notation

$$\mathbf{AU} = \mathbf{UD}$$

Symmetric Matrices

$\mathbf{A}$ symmetric with distinct eigenvalues λ_i.
The equations $\mathbf{A}\mathbf{u}_i = \lambda_i \mathbf{u}_i$ can be written in matrix form as:

$$\underbrace{\begin{bmatrix} a_{11} & & a_{1d} \\ \vdots & \dots & \vdots \\ a_{d1} & & a_{dd} \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_d \\ | & & | \end{bmatrix}}_{\mathbf{U}} = \underbrace{\begin{bmatrix} | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_d \\ | & & | \end{bmatrix}}_{\mathbf{U}} \underbrace{\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_d \end{bmatrix}}_{\mathbf{D}}$$

In matrix notation

$$\mathbf{AU} = \mathbf{UD}$$

Since $\mathbf{U}$ is an orthogonal matrix, $\mathbf{U}^\top = \mathbf{U}^{-1}$, thus

Symmetric Matrices

$\mathbf{A}$ symmetric with distinct eigenvalues λ_i.
The equations $\mathbf{A}\mathbf{u}_i = \lambda_i \mathbf{u}_i$ can be written in matrix form as:

$$\underbrace{\begin{bmatrix} a_{11} & & a_{1d} \\ \vdots & \dots & \vdots \\ a_{d1} & & a_{dd} \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_d \\ | & & | \end{bmatrix}}_{\mathbf{U}} = \underbrace{\begin{bmatrix} | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_d \\ | & & | \end{bmatrix}}_{\mathbf{U}} \underbrace{\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_d \end{bmatrix}}_{\mathbf{D}}$$

In matrix notation

$$\mathbf{A}\mathbf{U} = \mathbf{U}\mathbf{D}$$

Since $\mathbf{U}$ is an orthogonal matrix, $\mathbf{U}^\top = \mathbf{U}^{-1}$, thus

Spectral Decomposition of a Symmetric Matrix

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^\top$$

Quadratic Form

Let $\mathbf{A}$ be a symmetric matrix, then

Quadratic Form

Let $\mathbf{A}$ be a symmetric matrix, then

$$f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{x}$$

Quadratic Form

Let **A** be a symmetric matrix, then

$$f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{x}$$

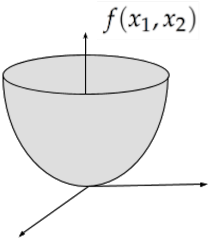
$$f(x_1, x_2) = [x_1\, x_2] \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1^2 + x_2^2$$

Quadratic Form

Let **A** be a symmetric matrix, then

$$f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{x}$$

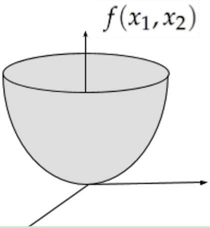
$$f(x_1, x_2) = [x_1\ x_2] \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1^2 + x_2^2$$



Quadratic Form

Let $\mathbf{A}$ be a symmetric matrix, then

$$f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{x}$$



$$f(x_1, x_2) = [x_1 \; x_2] \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1^2 + x_2^2$$

$$\max_{\|\mathbf{x}\|=1} \{f(\mathbf{x})\} = \mathbf{u}_1^\top \mathbf{A}\mathbf{u}_1 = \lambda_1$$

$$\min_{\|\mathbf{x}\|=1} \{f(\mathbf{x})\} = \mathbf{u}_d^\top \mathbf{A}\mathbf{u}_d = \lambda_d$$

$(\lambda_1, \mathbf{u}_1)$ and $(\lambda_d, \mathbf{u}_d)$ are the larger and smaller eigenpair.

Covariance Matrix

Let $\mathbf{x}_i = [x_{1i}, \ldots, x_{di}]^\top$, $\mathbf{x}_j = [x_{1j}, \ldots, x_{dj}]^\top$

Covariance Matrix

Let $\mathbf{x}_i = [x_{1i}, \ldots, x_{di}]^\top$, $\mathbf{x}_j = [x_{1j}, \ldots, x_{dj}]^\top$

The covariance between $\mathbf{x}_i$ and $\mathbf{x}_j$ is given by

$$cov(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{d-1} \sum_{s=1}^{d} (x_{si} - \overline{x}_i)(x_{sj} - \overline{x}_j)$$

where $\overline{x}_i = \frac{1}{d} \sum_s x_{si}$ and $\overline{x}_j = \frac{1}{d} \sum_s x_{sj}$

Covariance Matrix

Let $\mathbf{x}_i = [x_{1i}, \ldots, x_{di}]^\top$, $\mathbf{x}_j = [x_{1j}, \ldots, x_{dj}]^\top$

The covariance between $\mathbf{x}_i$ and $\mathbf{x}_j$ is given by

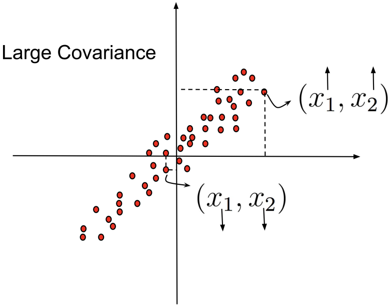
$$cov(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{d-1} \sum_{s=1}^{d} (x_{si} - \overline{x}_i)(x_{sj} - \overline{x}_j)$$

where $\overline{x}_i = \frac{1}{d} \sum_s x_{si}$ and $\overline{x}_j = \frac{1}{d} \sum_s x_{sj}$

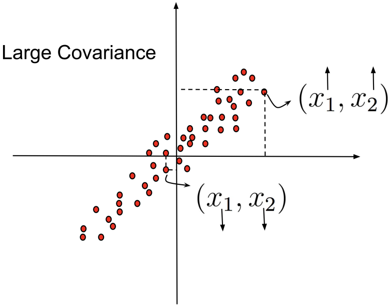
If we assume $\mathbf{x}_i$ and $\mathbf{x}_j$ centered, that is, $\overline{x}_i = 0$ and $\overline{x}_j = 0$

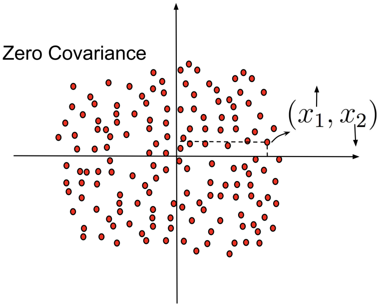
$$cov(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{d-1} \sum_s x_{si} x_{sj}$$

Covariance Matrix



Covariance Matrix





Covariance Matrix

Assuming $\mathbf{x}_i$, $i = 1, \ldots, n$ a centered set of data instances (points in $\mathbb{R}^d$) arranged in a data matrix $\mathbf{X}$:

$$\mathbf{X} = \begin{bmatrix} | & | & & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \ldots & \mathbf{x}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1n} \\ x_{21} & x_{22} & \ldots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{d1} & x_{d2} & \ldots & x_{dn} \end{bmatrix} \tag{1}$$

Covariance Matrix

Assuming $\mathbf{x}_i$, $i = 1, \ldots, n$ a centered set of data instances (points in $\mathbb{R}^d$) arranged in a data matrix $\mathbf{X}$:

$$\mathbf{X} = \begin{bmatrix} | & | & & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \ldots & \mathbf{x}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1n} \\ x_{21} & x_{22} & \ldots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{d1} & x_{d2} & \ldots & x_{dn} \end{bmatrix} \tag{1}$$

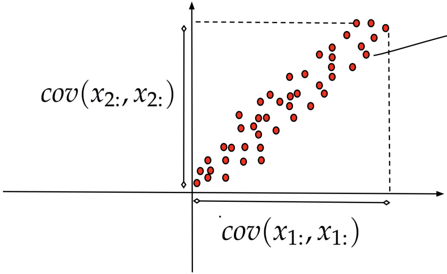
The covariance matrix of $\mathbf{X}$ is the symmetric matrix:

$$\frac{1}{n-1}\mathbf{X}\mathbf{X}^\top = \begin{bmatrix} \boxed{cov(x_{1:}, x_{1:})} & cov(x_{1:}, x_{2:}) & \ldots & cov(x_{1:}, x_{d:}) \\ cov(x_{2:}, x_{1:}) & \boxed{cov(x_{2:}, x_{2:})} & \ldots & cov(x_{2:}, x_{d:}) \\ \vdots & \vdots & \ddots & \vdots \\ cov(x_{d:}, x_{1:}) & cov(x_{d:}, x_{2:}) & \ldots & \boxed{cov(x_{d:}, x_{d:})} \end{bmatrix}$$

Variances are in the main diagonal

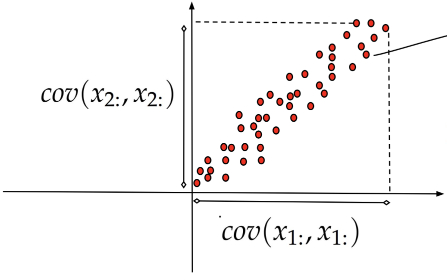
Principal Components: getting some intuition

Principal Components: getting some intuition



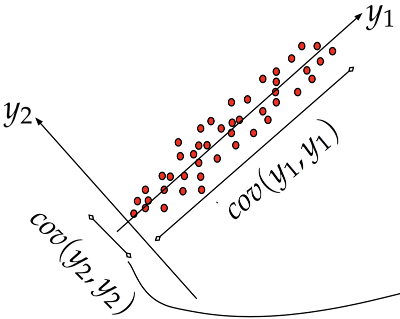
$$\mathbf{X}\mathbf{X}^\top = \begin{bmatrix} cov(x_{1:}, x_{1:}) & cov(x_{1:}, x_{2:}) \\ cov(x_{2:}, x_{1:}) & cov(x_{2:}, x_{2:}) \end{bmatrix}$$

Principal Components: getting some intuition



not zero

$$\mathbf{X}\mathbf{X}^\top = \begin{bmatrix} cov(x_{1:}, x_{1:}) & cov(x_{1:}, x_{2:}) \\ cov(x_{2:}, x_{1:}) & cov(x_{2:}, x_{2:}) \end{bmatrix}$$



$$\mathbf{Y}\mathbf{Y}^\top = \begin{bmatrix} cov(y_1, y_1) & 0 \\ 0 & cov(y_2, y_2) \end{bmatrix}$$

close to zero

Principal Components

The idea of PCA is to find a new basis to write the data so as to vanish the covariance between distinct attributes.

Principal Components

The idea of PCA is to find a new basis to write the data so as to vanish the covariance between distinct attributes.

Mathematically, we are looking for a change of basis matrix $\mathbf{P}$ such that

$$\mathbf{Y} = \mathbf{P}\mathbf{X} \Longrightarrow \mathbf{Y}\mathbf{Y}^\top = \mathbf{D}$$

where $\mathbf{D}$ is a diagonal matrix with diagonal elements corresponding to the variance of each coordinate (attribute).

Principal Components

The idea of PCA is to find a new basis to write the data so as to vanish the covariance between distinct attributes.

Mathematically, we are looking for a change of basis matrix $\mathbf{P}$ such that

$$\mathbf{Y} = \mathbf{P}\mathbf{X} \Longrightarrow \mathbf{Y}\mathbf{Y}^\top = \mathbf{D}$$

where $\mathbf{D}$ is a diagonal matrix with diagonal elements corresponding to the variance of each coordinate (attribute).

By fiding $\mathbf{P}$:

Principal Components

The idea of PCA is to find a new basis to write the data so as to vanish the covariance between distinct attributes.

Mathematically, we are looking for a change of basis matrix **P** such that

$$\mathbf{Y} = \mathbf{P}\mathbf{X} \Longrightarrow \mathbf{Y}\mathbf{Y}^\top = \mathbf{D}$$

where **D** is a diagonal matrix with diagonal elements corresponding to the variance of each coordinate (attribute).

By fiding **P**:

- the new attributes/coordinates will be decorrelated (redundancy removed)

Principal Components

The idea of PCA is to find a new basis to write the data so as to vanish the covariance between distinct attributes.

Mathematically, we are looking for a change of basis matrix $\mathbf{P}$ such that

$$\mathbf{Y} = \mathbf{P}\mathbf{X} \Longrightarrow \mathbf{Y}\mathbf{Y}^{\top} = \mathbf{D}$$

where $\mathbf{D}$ is a diagonal matrix with diagonal elements corresponding to the variance of each coordinate (attribute).

By fiding $\mathbf{P}$:

- the new attributes/coordinates will be decorrelated (redundancy removed)
- some coordinates will tend to be of low variance (noise related coordinates)

Principal Components

The idea of PCA is to find a new basis to write the data so as to vanish the covariance between distinct attributes.

Mathematically, we are looking for a change of basis matrix $\mathbf{P}$ such that

$$\mathbf{Y} = \mathbf{P}\mathbf{X} \Longrightarrow \mathbf{Y}\mathbf{Y}^\top = \mathbf{D}$$

where $\mathbf{D}$ is a diagonal matrix with diagonal elements corresponding to the variance of each coordinate (attribute).

By fiding $\mathbf{P}$:

- the new attributes/coordinates will be decorrelated (redundancy removed)
- some coordinates will tend to be of low variance (noise related coordinates)
- we will be able to reduce the dimension of the data without loosing relevant information.

Principal Components

$$\mathbf{Y} = \mathbf{P}\mathbf{X}$$

Principal Components

$$\mathbf{Y} = \mathbf{P}\mathbf{X}$$

$$\mathbf{Y}\mathbf{Y}^\top = (\mathbf{P}\mathbf{X})(\mathbf{P}\mathbf{X})^\top = \mathbf{P}\mathbf{X}\mathbf{X}^\top\mathbf{P}^\top$$

Principal Components

$$\mathbf{Y} = \mathbf{P}\mathbf{X}$$

$$\mathbf{Y}\mathbf{Y}^\top = (\mathbf{P}\mathbf{X})(\mathbf{P}\mathbf{X})^\top = \mathbf{P}\mathbf{X}\mathbf{X}^\top\mathbf{P}^\top$$

Reminder

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^\top$$
$$\downarrow$$
$$\mathbf{U}^\top\mathbf{A}\mathbf{U} = \mathbf{D}$$

Principal Components

$$\mathbf{Y} = \mathbf{P}\mathbf{X}$$

$$\mathbf{Y}\mathbf{Y}^\top = (\mathbf{P}\mathbf{X})(\mathbf{P}\mathbf{X})^\top = \mathbf{P}\mathbf{X}\mathbf{X}^\top\mathbf{P}^\top$$

Reminder

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^\top$$
$$\downarrow$$
$$\mathbf{U}^\top\mathbf{A}\mathbf{U} = \mathbf{D}$$

Eigenvectors of $\mathbf{X}\mathbf{X}^\top \rightarrow \mathbf{U}$

Principal Components

$$\mathbf{Y} = \mathbf{P}\mathbf{X}$$

$$\mathbf{Y}\mathbf{Y}^\top = (\mathbf{P}\mathbf{X})(\mathbf{P}\mathbf{X})^\top = \mathbf{P}\mathbf{X}\mathbf{X}^\top\mathbf{P}^\top$$

Reminder

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^\top$$
$$\downarrow$$
$$\mathbf{U}^\top\mathbf{A}\mathbf{U} = \mathbf{D}$$

Eigenvectors of $\mathbf{X}\mathbf{X}^\top \to \mathbf{U}$

$$\mathbf{P} = \mathbf{U}^\top$$

Principal Components

$$\mathbf{Y} = \mathbf{P}\mathbf{X}$$

$$\mathbf{Y}\mathbf{Y}^\top = (\mathbf{P}\mathbf{X})(\mathbf{P}\mathbf{X})^\top = \mathbf{P}\mathbf{X}\mathbf{X}^\top\mathbf{P}^\top$$

Reminder

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^\top$$
$$\downarrow$$
$$\mathbf{U}^\top\mathbf{A}\mathbf{U} = \mathbf{D}$$

Eigenvectors of $\mathbf{X}\mathbf{X}^\top \rightarrow \mathbf{U}$

$$\mathbf{P} = \mathbf{U}^\top$$

$$\mathbf{Y}\mathbf{Y}^\top = \mathbf{U}^\top\mathbf{X}\mathbf{X}^\top\mathbf{U}$$

Principal Components

$$\mathbf{Y} = \mathbf{P}\mathbf{X}$$

$$\mathbf{Y}\mathbf{Y}^\top = (\mathbf{P}\mathbf{X})(\mathbf{P}\mathbf{X})^\top = \mathbf{P}\mathbf{X}\mathbf{X}^\top\mathbf{P}^\top$$

Reminder

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^\top$$

$$\downarrow$$

$$\mathbf{U}^\top\mathbf{A}\mathbf{U} = \mathbf{D}$$

Eigenvectors of $\mathbf{X}\mathbf{X}^\top \rightarrow \mathbf{U}$

$$\mathbf{P} = \mathbf{U}^\top$$

$$\mathbf{Y}\mathbf{Y}^\top = \mathbf{U}^\top\mathbf{X}\mathbf{X}^\top\mathbf{U}$$

$$\mathbf{Y}\mathbf{Y}^\top = \mathbf{U}^\top\mathbf{X}\mathbf{X}^\top\mathbf{U} = \mathbf{D}$$

Principal Components

The coordinates of the data in the new basis is given by:

$$\mathbf{Y} = \mathbf{U}^\top \mathbf{X}$$

Principal Components

The coordinates of the data in the new basis is given by:

$$\mathbf{Y} = \mathbf{U}^\top \mathbf{X}$$

The diagonal matrix $\mathbf{D}$ in the decomposition $\mathbf{X}\mathbf{X}^\top = \mathbf{U}\mathbf{D}\mathbf{U}^\top$ contains the variances of each new coordinate.

Principal Components

The coordinates of the data in the new basis is given by:

$$\mathbf{Y} = \mathbf{U}^\top \mathbf{X}$$

The diagonal matrix $\mathbf{D}$ in the decomposition $\mathbf{X}\mathbf{X}^\top = \mathbf{U}\mathbf{D}\mathbf{U}^\top$ contains the variances of each new coordinate.

Moreover,

$$\mathbf{u}_1^\top \mathbf{X}\mathbf{X}^\top \mathbf{u}_1 = \lambda_1 \text{ (maximum of the quadratic form)}$$

$$\mathbf{u}_d^\top \mathbf{X}\mathbf{X}^\top \mathbf{u}_d = \lambda_d \text{ (minimum of the quadratic form)}$$

Principal Components

The coordinates of the data in the new basis is given by:

$$\mathbf{Y} = \mathbf{U}^\top \mathbf{X}$$

The diagonal matrix $\mathbf{D}$ in the decomposition $\mathbf{X}\mathbf{X}^\top = \mathbf{U}\mathbf{D}\mathbf{U}^\top$ contains the variances of each new coordinate.

Moreover,

$$\mathbf{u}_1^\top \mathbf{X}\mathbf{X}^\top \mathbf{u}_1 = \lambda_1 \text{ (maximum of the quadratic form)}$$

$$\mathbf{u}_d^\top \mathbf{X}\mathbf{X}^\top \mathbf{u}_d = \lambda_d \text{ (minimum of the quadratic form)}$$

$\mathbf{u}_1$ is the direction that maximizes the variance and $\mathbf{u}_d$ the direction that minimizes the variance.

Principal Components

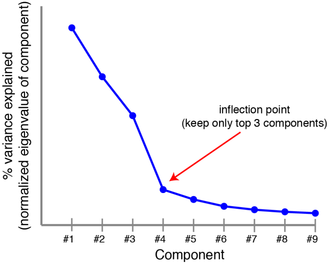
We can filter out low variance directions (corresponding to small λ_i), since they typically correspond to noise.

Principal Components

We can filter out low variance directions (corresponding to small λ_i), since they typically correspond to noise.

We can reconstruct "noise-free" data by $\hat{\mathbf{X}} = \mathbf{U}\hat{\mathbf{Y}}$

$$\hat{\mathbf{Y}} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ & & \vdots & \\ y_{k1} & y_{k2} & \cdots & y_{kn} \\ 0 & 0 & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$



$$T = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{d} \lambda_i}$$